

STUDI EKSPERIMENTAL : MENGURANGI BIAS PENGUKURAN UMUM BALANCED SCORECARD DALAM PENILAIAN KINERJA PADA MAHASISWA S1 PROGRAM MANAJEMEN PERHOTELAN

Jessica Octavia dan Juniarti
Akuntansi Bisnis Universitas Kristen Petra
Email: yunie@petra.ac.id

ABSTRAK

Penelitian Lipe dan Salterio (2000) menemukan adanya bias pengukuran umum. Bias pengukuran umum dapat mengurangi manfaat dari BSC. Beberapa penelitian sebelumnya melakukan berbagai pendekatan untuk mengurangi bias pengukuran umum (*debiasing*), Roberts, Albright & Hibbets (2004) dengan pendekatan *disaggregated/ mechanically aggregated* serta Dilla dan Steinbart (2005) dengan pendekatan pengetahuan mengenai BSC. Kedua penelitian sebelumnya berhasil mengurangi bias namun belum optimal maka dari itu dilakukan penelitian ini yang menggabungkan kedua pendekatan tersebut serta melihat apakah evaluasi kinerja menggunakan BSC mempengaruhi alokasi kompensasi. Metodologi penelitian yang dilakukan adalah penelitian eksperimen laboratorium. Pengujian hipotesis penelitian ini menggunakan uji *repeated measure ANOVA*, *independent sample t test* serta analisis regresi.

Hasil penelitian menunjukkan bahwa pendekatan *disaggregated/ mechanically aggregated* berpengaruh signifikan dalam mengurangi bias pengukuran umum. Ditemukan juga bahwa pendekatan pengetahuan mengenai BSC berpengaruh signifikan dalam mengurangi bias pengukuran umum, serta evaluasi kinerja menggunakan BSC berpengaruh signifikan terhadap alokasi kompensasi.

Kata kunci: Balanced Scorecard, pengukuran umum, pengukuran unik, *debiasing*, bias pengukuran umum, *disaggregated/ mechanically aggregated*, pengetahuan, alokasi kompensasi.

ABSTRACT

Lipe & Salterio (2000) found a common measure bias. Common measurement bias may reduce the benefits of BSC. Several previous studies perform a variety of approaches to reduce the common measurement bias (*debiasing*), Roberts, Albright & Hibbets (2004) with a *disaggregated/ mechanically aggregated* approach and Dilla & Steinbart (2005) with the knowledge of BSC approach. Both studies successfully reduces bias but not optimal and therefore this research conducted by combining both approaches and seen if the performance evaluation by using BSC affected the allocation of compensation. The methodology used was laboratory experimental research. The hypothesis of this study tested by using a *repeated measure ANOVA* test, *independent sample t test* and regression analysis.

The results showed that the *disaggregated/ aggregated mechanically* approach significant by affected in reducing the general measurement bias. It was also found that the knowledge approach of the BSC approach significant by affected in reducing common measure bias, and performance evaluation by using the BSC significantly influenced the allocation of compensation.

Keywords: Balanced Scorecard, common measure, unique measure, *debiasing*, common measure bias, *disaggregated/ mechanically aggregated*, knowledge, allocation of compensation.

PENDAHULUAN

Balanced scorecard (BSC) menjadi salah satu alat pengukuran kinerja yang paling banyak digunakan oleh perusahaan hal ini didukung oleh survei yang dilakukan oleh Rigby & Bilodeau (2013). BSC tidak hanya digunakan sebagai alat pengukuran kinerja, tetapi juga digunakan sebagai alat untuk mengimplementasi dan memonitor strategi perusahaan (Kaplan & Norton, 1996a, 2001a, 2001b).

Implementasi BSC didalam perusahaan membutuhkan biaya yang relatif mahal (Ittner, Larcker & Meyer, 2003; Lippe & Salterio, 2000; Lindberg & Schonfeldt, 2008) maka penggunaan BSC harus dilakukan secara maksimal. Menurut Lipe & Salterio (2000) manfaat yang diperoleh dalam mengadopsi BSC tergantung pada sejauh mana BSC meningkatkan keputusan manajer. Salah satu keputusan manajer adalah keputusan pemberian kompensasi. Kaplan & Norton (1996, 2001) sendiri menyarankan bagi perusahaan untuk menghubungkan kompensasi dengan BSC untuk meningkatkan manfaat BSC.

Hasil pengukuran kinerja akan menjadi sumber pertimbangan untuk pengambilan kompensasi bagi karyawan, namun karena adanya bias pengukuran umum pengambilan kompensasi menjadi tidak tepat. Hal ini dapat menyebabkan karyawan merasa tidak adil karena hasil kinerjanya dinilai secara tidak komprehensif (Liedka, Church & Ray, 2008; Gibbs, Merchant, Van der Stede & Vargus, 2005). Serta penurunan kinerja perusahaan karena karyawan akan mengurangi usaha mereka terhadap aktivitas-aktivitas yang dianggap tidak dievaluasi (Holstrom & Milgrom, 1991; Malina & Selto, 2001).

Penelitian yang dilakukan Lipe & Salterio (2000) menemukan adanya dominasi pengukuran umum yang mengakibatkan munculnya bias pengukuran umum. Terdapat banyak penelitian yang dilakukan untuk mengurangi bias pengukuran umum, salah satunya penelitian yang dilakukan oleh Roberts, Albright & Hibbets (2004) yang menggunakan pendekatan *disaggregated / mechanically aggregated* untuk mengatasi bias pengukuran umum. Pendekatan *disaggregated / mechanically aggregated* berhasil mengurangi bias pengukuran umum namun bias pengukuran umum masih tetap ada.

Penelitian lainnya yang dilakukan Dilla & Steinbart (2005) menggunakan pendekatan pengetahuan mengenai BSC kepada para partisipan, dimana menurut Dilla & Steinbart pengetahuan merupakan suatu cara untuk memperoleh informasi yang lebih banyak. Dengan

pengetahuan yang lebih akan menghasilkan penilaian yang lebih baik (Bonner, 1990; Bonner & Lewis, 1990). Hasil dari penelitian ini menemukan bias pengukuran umum berkurang namun tidak maksimal.

Penelitian yang dilakukan Roberts, Albright & Hibbets (2004) serta Dilla & Steinbart (2005) berhasil mengurangi bias pengukuran umum namun tidak secara maksimal. Penyebab tidak maksimalnya pengurangan bias pengukuran umum karena kedua pendekatan tersebut dilakukan secara parsial atau terpisah.

Berdasarkan latar belakang permasalahan yang telah diungkapkan, peneliti kemudian tertarik untuk mengevaluasi kembali BSC dengan studi eksperimen menggunakan cara yang menggabungkan pendekatan yang dilakukan oleh Roberts, Albright & Hibbets (2004) dengan pendekatan *disaggregated / mechanically aggregated* BSC dan pendekatan yang dilakukan oleh Dilla & Steinbart (2005) yaitu memberikan pengetahuan terlebih dahulu mengenai BSC kepada partisipan. Kedua pendekatan dari peneliti sebelumnya akan digunakan pada studi eksperimen peneliti pada saat ini untuk menilai apakah bias pengukuran umum dapat berkurang pada mahasiswa S1 Program Manajemen Perhotelan.

Pengertian Pengukuran Umum dan Unik

Pengukuran BSC selain terdiri dari ukuran keuangan dan non keuangan (Kaplan & Norton, 1992). Ditinjau dari strategi perusahaan terdapat ukuran umum dan unik (Kaplan & Norton, 1996, 2000, 2001; Lipe & Salterio, 2000; Libby, Salterio & Webb, 2002; Roberts, Albright & Hibbets, 2004; Dilla & Steinbart, 2005).

Pengukuran umum adalah pengukuran kinerja perusahaan yang diterapkan diseluruh unit bisnis perusahaan (Kaskey, 2008). Setiap unit bisnis didalam perusahaan merupakan bagian dari perusahaan, maka dari itu beberapa ukuran berlaku secara umum bagi seluruh unit bisnis perusahaan, hal ini diidentifikasi sebagai pengukuran umum (Gagne, Hollister & Tully., 2006; Kaskey, 2008; Bawono, Halim & Lord, 2012). Contoh pengukuran umum pada perspektif keuangan yaitu *return on sales & sales growth*, pada perspektif pelanggan yaitu *repeat sales & customer satisfaction rating*, pada perspektif proses bisnis internal yaitu *returns to suppliers & average markdowns*, pada perspektif pembelajaran dan pertumbuhan yaitu *hours of employee training per employee & employee suggestion per employee* (Lipe & Salterio, 2000).

Pengukuran unik adalah pengukuran kinerja masing-masing unit bisnis perusahaan (Lipe & Salterio, 2000). Setiap unit bisnis memiliki strategi

dan tujuan spesifik yang berbeda-beda sehingga beberapa ukuran BSC akan bervariasi sesuai dengan strategi dan tujuan spesifik unit bisnis (Kaplan & Norton, 1993; Lipe & Salterio, 2000; Gagne, Hollister & Tully, 2006). Menurut Kaplan & Norton (1996) kekuatan penting dari BSC adalah bahwa setiap unit bisnis dalam organisasi akan memiliki *scorecard* sendiri yang secara khusus dirancang untuk unit tersebut. Contoh pengukuran unik pada perspektif keuangan yaitu *new store sales & revenue per sales visit*, pada perspektif pelanggan yaitu *captured customers & referrals*, pada perspektif proses bisnis internal yaitu *orders filled within one week & catalog orders filled with errors*, pada perspektif pembelajaran dan pertumbuhan yaitu *average tenure of sales personnel & stores computerizing* (Lipe & Salterio, 2000).

Pengertian Bias Pengukuran Umum

Bias pengukuran umum sering dijelaskan sebagai ketidakmampuan pengambil keputusan untuk menyertakan informasi unik dalam evaluasi kinerja, karena informasi ini membutuhkan usaha kognitif yang lebih untuk dapat diproses (Lipe & Salterio, 2000, 2005; Roberts, Albright & Hibbets, 2004; Libby, Salterio & Webb, 2004).

Individu memiliki keterbatasan kognitif yaitu keterbatasan menyimpan informasi dalam memori/*bounded rationality* (Tversky & Kahneman, 1973, 1974; Miqdad, 2012) dalam pengambilan keputusan, termasuk keputusan bisnis (Bone & Sholihin, 2012). Penelitian Slovic & MacPhillamy (1974) menemukan informasi umum memiliki dampak yang lebih besar karena lebih mudah digunakan dalam membandingkan kandidat yang dinilai (dalam Roberts, Albright & Hibbets, 2004). Ukuran unik tidak memiliki pembandingan pada saat penilaian, sementara ukuran umum menyediakan informasi yang secara langsung dapat digunakan untuk membandingkan hasil kinerja (Roberts, Albright & Hibbets, 2004; Dilla & Steinbart, 2005). Hal ini menunjukkan para pengambil keputusan cenderung menggunakan ukuran umum untuk menyederhanakan tugas penilaiannya (Payne, Bettman & Johnson, 1993).

Hasil penelitian Lipe & Salterio (2000) menemukan partisipan yang mengevaluasi kinerja perusahaan cenderung menggunakan ukuran umum dan mengabaikan ukuran unik yang ada. Dominasi pengukuran umum akan mengakibatkan munculnya bias pengukuran umum.

Dampak adanya bias pengukuran umum adalah akan mengurangi manfaat dari BSC (Lipe & Salterio, 2000; Libby, Salterio & Webb, 2002;

Roberts, Albright & Hibbets, 2004; Dilla & Steinbart, 2005; Aryani, 2009; Humphreys & Trotman, 2011). Selain itu bias pengukuran umum dapat menyebabkan keputusan pemberian kompensasi yang tidak tepat dan adil (Liedtka, Church & Ray, 2008; Bone & Sholihin, 2012; Krumwiede, Swain, Thornock & Eggett, 2012). Karyawan merasa tidak adil karena hasil kinerjanya tidak dinilai secara komprehensif (Liedtka, Church & Ray, 2008).

Debiasing (Mengurangi Bias Pengukuran Umum)

Bazerman & Moore (2009) mendefinisikan *debiasing* adalah suatu prosedur untuk mengurangi atau mengeliminasi bias dari strategi kognitif pengambil keputusan. Menurut Serfas (2011) terdapat tiga level teknik untuk mengeliminasi bias. Pertama adalah *knowledge*, *experience* dan *expertise*, kedua adalah *incentive* dan *accountability*, dan yang ketiga adalah *training* dan *tools*. Untuk mendukung keberhasilan *debiasing* dibutuhkan waktu dan usaha dari para pengambil keputusan (Serfas, 2010). Ketika usaha yang diberikan tidak cukup untuk meningkatkan kualitas penilaian dapat menggunakan alat bantu keputusan.

Menurut Einhorn (1972) menemukan adanya peningkatan akurasi penilaian manusia dapat terjadi ketika informasi mengenai keputusan diubah menjadi bentuk kuantitatif dan hasilnya digeneralisasi menggunakan *mechanical combination rule* (dalam Roberts, Albright & Hibbets, 2004). Bowman (1963) menyarankan untuk mengkombinasikan manusia dan model dengan menggunakan "*clinical synthesis*", dimana individu menggunakan output dari model sebagai *input* untuk keputusan akhir individu (dalam Roberts, Albright & Hibbets, 2004).

Aplikasi dari temuan Einhorn (1972) dan Bowman (1963) menyarankan BSC meliputi dua langkah : (1) *mendisagregatkan* keputusan menjadi beberapa keputusan yang lebih kecil dan (2) *mengaggregatkan* keputusan yang lebih kecil ke dalam nilai total berdasarkan *predetermined weights* (Roberts, Albright & Hibbets, 2004).

Dapat diartikan pendekatan *disaggregated /mechanically aggregated* merupakan alat untuk mengurangi bias dengan menyederhanakan informasi yang perlu di proses (Roberts, Albright & Hibbets, 2004). Pendekatan *disaggregated /mechanically aggregated* dalam hal ini juga diartikan sebagai pendekatan yang tidak membandingkan kinerja satu unit dengan unit lainnya tetapi dengan melakukan penilaian secara individual untuk masing-masing unit (Grevinga, 2013).

Pendekatan *disaggregated / mechanically aggregated* akan menguntungkan penilaian yang syaratnya sulit (Roberts, Albright & Hibbets, 2004). Pendekatan *disaggregated / mechanically aggregated* akan menurunkan dan meningkatkan tuntutan tugas bagi manajer, tuntutan kognitif berkurang karena jumlah informasi yang harus dipertimbangkan untuk mengevaluasi setiap individu berkurang daripada informasi seluruh BSC, namun total waktu dan upaya yang diperlukan meningkat karena jumlah evaluasi dan perhitungan meningkat (Roberts, Albright & Hibbets, 2004). Selain itu dengan pendekatan *disaggregated / mechanically aggregated* mampu mengurangi bias pengukuran umum karena BSC yang dipisahkan (*disaggregated*) lebih memberikan perhatian pada ukuran unik dalam divisi tertentu tidak hanya pada ukuran umum saja (Grevinga, 2013).

Pendekatan lainnya yang dilakukan untuk mengurangi bias pengukuran umum adalah dengan pengetahuan (Dilla & Steinbart, 2005). Hal ini sejalan dengan hasil penelitian penilaian audit menunjukkan bahwa pengetahuan yang lebih luas akan menghasilkan penilaian yang lebih baik di berbagai macam tugas audit (Bonner, 1990; Bonner & Lewis, 1990; Libby & Tan, 1994).

Hal tersebut relevan dengan keputusan penilaian menggunakan BSC para pengambil keputusan yang lebih memahami teori dan struktur BSC akan lebih mudah untuk mengabungkan kedua ukuran, baik ukuran umum dan unik ketika membandingkan dan mengevaluasi kinerja masing-masing unit bisnis (Dilla & Steinbart, 2005). Didukung pula oleh penelitian Niven (2006) bahwa pelatihan dan pelajaran tentang teori dan rancangan dari BSC sangat penting untuk kesuksesan penerapan BSC itu sendiri.

Pelatihan formal merupakan salah satu cara untuk memperoleh pengetahuan. Salah satu metode yang digunakan untuk memperoleh pengetahuan yang ada adalah melalui pembelajaran yang didapatkan pada perkuliahan. Melalui pembelajaran dari perkuliahan tersebut maka akan diperoleh pengetahuan mengenai struktur BSC dan contoh-contoh BSC untuk berbagai jenis organisasi.

Pengukuran BSC & Alokasi Kompensasi

BSC pada awalnya diperkenalkan sebagai alat pengukuran kinerja (Kaplan & Norton, 1992), kemudian berkembang bukan hanya sebagai alat pengukuran kinerja namun juga sebagai alat implementasi dan memonitor strategi (Kaplan & Norton, 1996, 2001). Beberapa pendukung BSC mengatakan bahwa manfaat dari BSC akan

meningkat jika dihubungkan dengan keputusan pemberian kompensasi (Kaplan & Norton, 1996, 2001; Frigo & Krumwiede, 2000; Niven, 2006).

Pada penelitian Lipe & Salterio (2000) yang menemukan adanya bias pengukuran umum dimana para partisipan yang bertindak sebagai para manajer hanya fokus pada pengukuran umum dan mengabaikan pengukuran unik ketika mengevaluasi kinerja karyawan. Dengan adanya bias pengukuran umum hasil keputusan manajer menjadi tidak tepat, sehingga pemberian kompensasi kepada karyawan juga menjadi tidak tepat. Keputusan kompensasi dipengaruhi oleh nilai evaluasi kinerja dari manajer (Grevinga, 2013). Pemberian kompensasi yang tidak tepat akan menyebabkan karyawan merasa tidak adil karena kinerja para karyawan tidak dinilai secara komprehensif (Liedtka, Church & Ray, 2008) dan menciptakan ketidakpuasan terhadap BSC (Dilla & Steinbart, 2005). Ketidak puasan terhadap BSC dapat menyebabkan perusahaan berhenti menghubungkan BSC dengan keputusan kompensasi (Ittner, Larcker & Meyer, 2003).

Hal tersebut akan berdampak pada kinerja karyawan, karena karyawan akan mengurangi usaha mereka terhadap aktivitas-aktivitas yang dianggap tidak dievaluasi (Holstrom & Milgrom, 1991; Malina & Selto, 2001). Dampak lainnya adalah mengurangi semangat karyawan untuk berusaha (Gibbs, Merchant, Van der Stede & Vargus, 2004; Liedtka, Church & Ray, 2005). Dengan adanya pengurangan usaha karyawan terhadap aktivitas-aktivitas yang dianggap tidak dievaluasi dimana dalam hal ini adalah pengukuran unik akan menghambat perusahaan mencapai tujuannya, karena semua ukuran dalam BSC dibuat untuk mencapai tujuan perusahaan (Kaplan & Norton, 1996).

Pengaruh pendekatan *Disaggregated/ Mechanically Aggregated* terhadap bias pengukuran umum

Penelitian Lipe & Salterio (2000) menemukan partisipan eksperimen cenderung menggunakan ukuran umum dibandingkan dengan ukuran unik dalam melakukan penilaian kinerja perusahaan. Dominasi pengukuran umum menyebabkan munculnya bias pengukuran umum.

Bias pengukuran umum sering dijelaskan sebagai ketidakmampuan pengambil keputusan untuk menyertakan informasi unik dalam evaluasi kinerja, karena informasi ini membutuhkan usaha kognitif yang lebih untuk dapat diproses (Lipe & Salterio, 2000, 2005; Roberts, Albright & Hibbets, 2004; Libby, Salterio & Webb, 2004).

Dampak adanya bias pengukuran umum adalah akan mengurangi manfaat dari BSC (Lipe &

Salterio, 2000; Libby, Salterio & Webb, 2002; Roberts, Albright & Hibbets, 2004; Dilla & Steinbart, 2005; Aryani, 2009). Selain itu bias pengukuran umum dapat menyebabkan keputusan pemberian kompensasi yang tidak tepat & adil. Karyawan merasa tidak adil karena hasil kinerjanya tidak dinilai secara komprehensif (Liedtka, Church & Ray, 2008)

Terdapat dua pendekatan untuk mengurangi bias pengukuran umum yang ada. Pertama dengan menggunakan pendekatan *disaggregated /mechanically aggregated* yang digunakan pada penelitian Roberts, Albright & Hibbets (2004). Pendekatan ini mengurangi bias dengan menyederhanakan informasi yang perlu diproses (Roberts, Albright & Hibbets, 2004), meliputi dua langkah: (1) *mendisagregatkan* keputusan evaluasi menjadi beberapa keputusan yang lebih kecil dan (2) *mengagregatkan* keputusan yang lebih kecil ke dalam nilai total berdasarkan *predetermined weights* (Roberts, Albright & Hibbets, 2004).

Pendekatan *disaggregated /mechanically aggregated* akan menguntungkan penilaian yang syaratnya sulit (Roberts, Albright & Hibbets, 2004). Pendekatan *disaggregated /mechanically aggregated* akan menurunkan dan meningkatkan tuntutan tugas bagi manajer, tuntutan kognitif berkurang karena jumlah informasi yang harus dipertimbangkan untuk mengevaluasi setiap individu berkurang daripada informasi seluruh BSC namun, total waktu dan upaya yang diperlukan meningkat karena jumlah evaluasi dan perhitungan meningkat (Roberts, Albright & Hibbets, 2004). Selain itu dengan pendekatan *disaggregated /mechanically aggregated* mampu mengurangi bias pengukuran umum karena BSC yang dipisahkan (*disaggregated*) lebih memberikan perhatian pada ukuran unik dalam divisi tertentu tidak hanya pada ukuran umum saja (Grevinga, 2013). Dengan demikian, maka sebuah hipotesis yang muncul adalah :

H1 : Pendekatan *disaggregated /mechanically aggregated* berpengaruh dalam mengurangi bias pengukuran umum dalam evaluasi kinerja menggunakan BSC.

Pengaruh pendekatan pengetahuan terhadap bias pengukuran umum

Pendekatan kedua untuk mengurangi bias pengukuran umum adalah dengan pendekatan pengetahuan yang digunakan pada penelitian Dilla & Steinbart (2005). Pengetahuan dapat diperoleh melalui pelatihan formal, seperti melalui perkuliahan. Keputusan penilaian menggunakan BSC oleh para pengambil keputusan yang lebih memahami teori dan struktur BSC akan lebih mudah untuk mengabungkan kedua pengukuran,

baik pengukuran umum dan unik ketika membandingkan dan mengevaluasi kinerja masing-masing unit bisnis (Dilla & Steinbart, 2005).

Pengambil keputusan yang memiliki pengetahuan mengenai tugas yang dilakukan akan memberikan penilaian dengan pertimbangan yang berbeda dengan pengambil keputusan yang kurang memiliki pengetahuan, khususnya untuk tugas yang rumit (Bonner, 1990). Penelitian Niven (2006) mengatakan bahwa pelatihan dan pelajaran tentang teori dan rancangan dari BSC sangat penting untuk kesuksesan penerapan BSC itu sendiri. Dengan demikian, maka sebuah hipotesis yang muncul adalah :

H2 : Pendekatan pengetahuan berpengaruh dalam mengurangi bias pengukuran umum dalam evaluasi kinerja menggunakan BSC.

Pengaruh penilaian kinerja terhadap alokasi kompensasi

Beberapa pendukung BSC mengatakan bahwa manfaat dari BSC akan meningkat jika dihubungkan dengan keputusan pemberian kompensasi (Frigo & Krumwiede, 2000; Kaplan & Norton, 1996, 2001; Niven, 2006).

Pemberian kompensasi yang tidak tepat akan menyebabkan karyawan merasa tidak adil karena kinerja para karyawan tidak dinilai secara komprehensif (Liedtka, Church & Ray, 2008) dan menciptakan ketidakpuasan terhadap BSC (Dilla & Steinbart, 2005).

Hal tersebut akan berdampak pada kinerja karyawan, karena karyawan akan mengurangi usaha mereka terhadap aktivitas-aktivitas yang dianggap tidak dievaluasi (Holstrom & Milgrom, 1991; Malina & Selto, 2001). Dampak lainnya adalah mengurangi semangat karyawan untuk berusaha (Gibbs, Merchant, Van der Stede & Vargus, 2004; Liedtka, Church & Ray, 2005). Jika usaha dan motivasi karyawan berkurang hal ini akan menyebabkan pada penurunan kinerja perusahaan. Dengan demikian, maka hipotesis yang muncul adalah :

H3 : Evaluasi kinerja menggunakan BSC akan mempengaruhi alokasi kompensasi.

METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah penelitian eksperimen laboratorium. Penelitian eksperimen laboratorium merupakan penelitian yang menguji hubungan sebab akibat pada lingkungan yang *artificial / buatan* (Indrianto & Supomo, 1999). Peneliti terlibat dalam pembuatan setting buatan dan

melakukan manipulasi terhadap variable tertentu (Indrianto & Supomo, 1999).

Peneliti menggunakan penelitian eksperimen laboratorium karena peneliti dapat mengontrol dan mengendalikan variabel eksperimen, sehingga hasil yang didapat lebih akurat serta validitas internal yang tinggi.

Desain Eksperimen

Desain eksperimen pada penelitian ini diadopsi dari eksperimen Lipe & Salterio (2000), Roberts, Albright & Hibbets (2004) dan Dilla & Steinbart (2005). Studi kasus ini mengenai perusahaan WCS yang bergerak dibidang retail yang khusus menjual pakaian wanita dan memiliki dua divisi, yaitu divisi RadWear dan WorkWear.

Penelitian ini menggunakan desain eksperimen 2 x 2 x 2 dengan dua level *between subject* (umum & unik) dilengkapi dengan satu level *within subject* (divisi). Faktor pertama *between subject*, yaitu umum menunjukkan apakah kinerja RadWear lebih baik pada pengukuran umum dibandingkan WorkWear atau sebaliknya. Faktor kedua *between subject*, yaitu unik menunjukkan apakah kinerja RadWear lebih baik pada pengukuran unik dibandingkan WorkWear atau sebaliknya. Setiap partisipan mengevaluasi kinerja manajer kedua divisi, dimana divisi merupakan *within subject*.

Partisipan penelitian eksperimen ini dibagi menjadi dua kelompok, yaitu kelompok partisipan yang diberikan pengetahuan mengenai BSC (kelompok *treatment*) dan kelompok partisipan yang tidak diberikan pengetahuan mengenai BSC (kelompok *non-treatment*). Kedua kelompok partisipan melakukan evaluasi kinerja BSC dengan pendekatan *disaggregated /mechanically aggregated*.

Partisipan diminta berperan sebagai senior eksekutif perusahaan yang akan menilai kinerja dua divisi, yaitu RadWear yang khusus untuk pakaian remaja dan WorkWear yang khusus untuk seragam kerja wanita. Partisipan akan mendapatkan studi kasus eksperimen berisi misi WCS dan strategi kedua divisi beserta dengan instruksi pengerjaan kasus. Peneliti akan memberikan penjelasan mengenai latar belakang perusahaan dan memberikan waktu 10 menit kepada partisipan untuk membaca studi kasus yang telah diberikan.

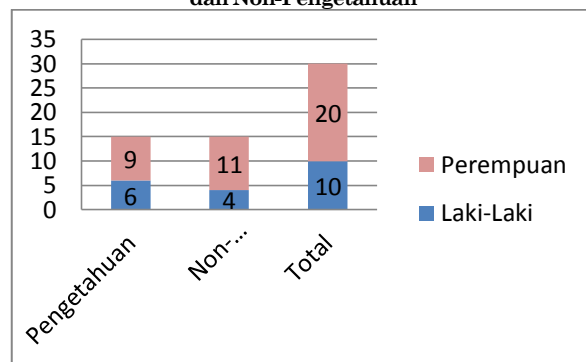
Langkah selanjutnya partisipan akan mendapatkan BSC divisi RadWear terlebih

dahulu. Dimana setiap BSC divisi memiliki 4 perspektif yang terdiri dari dua ukuran umum dan dua ukuran unik untuk setiap perspektif, dengan total 16 ukuran terpisah. BSC berisi *pre-determined weight* (bobot) untuk masing-masing ukuran dan persentase target kinerja divisi serta kinerja aktual divisi untuk setiap ukuran BSC.

HASIL PENELITIAN DAN PEMBAHASAN

Profil Partisipan

Tabel 1
Jenis Kelamin Partisipan pada Kelompok Pengetahuan dan Non-Pengetahuan



Uji Hipotesis 1

a. Uji ANOVA Pada Kelompok Pengetahuan

Tabel 2
Hasil Uji ANOVA pada Kelompok Pengetahuan

Tests of Between-Subjects Effects^a

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	37079.385	1	37079.385	29151.360	.000
Common	173.208	1	173.208	136.174	.000
Unique	70.856	1	70.856	55.706	.000
Common * Unique	6821.136	1	6821.136	5362.694	.000
Error	71.230	56	1.272		

a. Kode = Pengetahuan

Tests of Within-Subjects Contrasts^a

Measure: MEASURE_1

Source	division	Type III Sum of Squares	df	Mean Square	F	Sig.
division	Linear	70.856	1	70.856	55.706	.000
division * Common	Linear	6821.136	1	6821.136	5362.694	.000
division * Unique	Linear	37079.385	1	37079.385	29151.360	.000
division * Common * Unique	Linear	173.208	1	173.208	136.174	.000
Error(division)	Linear	71.230	56	1.272		

a. Kode = Pengetahuan

Berdasarkan data ini dapat disimpulkan bahwa nilai signifikansi pengukuran umum dan pengukuran unik dengan divisi pada kelompok pengetahuan < 0.05.

b. Uji ANOVA Pada Kelompok Non-Pengetahuan

Tabel 3
Hasil Uji ANOVA pada Kelompok Non-Pengetahuan

Tests of Between-Subjects Effects^a

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	35488.073	1	35488.073	31074.135	.000
Common	163.707	1	163.707	143.345	.000
Unique	16.703	1	16.703	14.625	.000
Common * Unique	6370.296	1	6370.296	5577.970	.000
Error	63.955	56	1.142		

a. Kode = Non Pengetahuan

Tests of Within-Subjects Contrasts^a

Measure: MEASURE_1

Source	division	Type III Sum of Squares	df	Mean Square	F	Sig.
division	Linear	16.703	1	16.703	14.625	.000
division * Common	Linear	6370.296	1	6370.296	5577.970	.000
division * Unique	Linear	35488.073	1	35488.073	31074.135	.000
division * Common * Unique	Linear	163.707	1	163.707	143.345	.000
Error(division)	Linear	63.955	56	1.142		

a. Kode = Non Pengetahuan

Berdasarkan data ini dapat disimpulkan bahwa nilai signifikansi pengukuran umum dan pengukuran unik dengan divisi pada kelompok non-pengetahuan < 0.05.

c. Uji ANOVA Pada Seluruh Kelompok

Tabel 4
Hasil Uji ANOVA pada Seluruh Kelompok

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	72558.733	1	72558.733	53539.766	.000
Common	336.848	1	336.848	248.554	.000
Unique	78.181	1	78.181	57.689	.000
Common * Unique	13187.579	1	13187.579	9730.874	.000
Error	157.207	116	1.355		

Tests of Within-Subjects Contrasts^a

Measure: MEASURE_1

Source	division	Type III Sum of Squares	df	Mean Square	F	Sig.
division	Linear	78.181	1	78.181	57.689	.000
division * Common	Linear	13187.579	1	13187.579	9730.874	.000
division * Unique	Linear	72558.733	1	72558.733	53539.766	.000
division * Common * Unique	Linear	336.848	1	336.848	248.554	.000
Error(division)	Linear	157.207	116	1.355		

Berdasarkan data ini dapat disimpulkan bahwa nilai signifikansi pengukuran umum dan pengukuran unik dengan divisi pada seluruh kelompok < 0.05.

Uji Hipotesis 2

Tabel 5
Hasil Independent Sample T-Test

Group Statistics

Kode	N	Mean	Std. Deviation	Std. Error Mean
Common Pengetahuan	15	40.1553	1.40344	.36237
Non Pengetahuan	15	39.6437	1.58287	.40869
Unique Pengetahuan	15	100.4707	2.65535	.68561
Non Pengetahuan	15	97.9317	3.47143	.89632

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Common	Equal variances assumed	.748	.395	.937	28	.357	.51167	.54621	-.60719	1.63052
	Equal variances not assumed			.937	27.604	.357	.51167	.54621	-.60791	1.63124
Unique	Equal variances assumed	.427	.519	2.250	28	.032	2.53900	1.12847	.22743	4.85057
	Equal variances not assumed			2.250	26.205	.033	2.53900	1.12847	.22027	4.85773

Berdasarkan data diatas dapat disimpulkan bahwa nilai signifikansi pengukuran umum pada kelompok pengetahuan dan non-pengetahuan memiliki nilai yang lebih besar dari 0.05. Sedangkan nilai signifikansi pengukuran unik pada kelompok pengetahuan dan non-pengetahuan memiliki nilai yang lebih kecil dari 0.05.

Uji Hipotesis 3

Tabel 6
Hasil Analisis Regresi

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
1 (Constant)	1.068	.082		13.101	.000		
Rc	.016	.004	.247	4.271	.000	.536	1.867
Ru	.016	.002	.604	9.777	.000	.470	2.129
Wc	-.020	.003	-.407	-6.910	.000	.517	1.935
Wu	-.016	.002	-.583	-10.677	.000	.601	1.685

a. Dependent Variable: Y_Rasio

Berdasarkan tabel diatas dapat di katakan nilai signifikansi RC, RU, WC, WU dibawah 0.05.

KESIMPULAN

Dari hasil penelitian yang dikemukakan diatas, maka temuan dari penelitian ini adalah sebagai berikut :

Pada hipotesis 1 pengujian data menggunakan uji *repeated measures* ANOVA menemukan bahwa pengukuran umum dan pengukuran unik digunakan secara proporsional, ditunjukkan dari nilai signifikansi pengukuran umum dan unik pada kedua divisi dibawah 0.05. Interaksi pengukuran umum dengan divisi menunjukkan nilai F = 9730.874, sedangkan interaksi pengukuran unik dengan divisi menunjukkan nilai F = 53539.766. Dilihat dai F hitung pengukuran unik yang lebih besar dapat dikatakan pengukuran unik berpengaruh

terhadap evaluasi kinerja menggunakan BSC. Oleh karena itu berdasarkan hasil penelitian ini maka **H1 di terima**, pendekatan *disaggregated/mechanically aggregated* dapat mengurangi bias pengukuran umum dalam evaluasi kinerja menggunakan BSC.

Pada hipotesis 2, hasil uji *independent sample t-test* menunjukan nilai signifikansi pengukuran umum lebih besar dari 0.05 sehingga dapat dikatakan tidak terdapat perbedaan yang signifikan pada pengukuran umum dalam kedua kelompok eksperimen. Sementara nilai signifikansi pengukuran unik lebih kecil dari 0.05 sehingga dapat dikatakan terdapat perbedaan yang signifikan pada pengukuran unik dalam kedua kelompok eksperimen. Berdasarkan hasil penelitian ini dapat dikatakan bahwa pendekatan pengetahuan mengenai BSC berhasil mengurangi bias pengukuran umum BSC dengan kata lain **H2 diterima**.

Pada hipotesis 3, dilihat dari nilai signifikansi analisis regresi menunjukan bahwa variabel RC, RU, WC, dan WU berpengaruh signifikan terhadap alokasi kompensasi. Dilihat dari nilai R Square 95,9% alokasi kompensasi dipengaruhi oleh RC, RU, WC, dan WU, sedangkan sisanya yaitu 4,1% dipengaruhi oleh faktor lain. Berdasarkan hasil tersebut dapat dikatakan **H3 diterima**, yaitu BSC mempengaruhi alokasi kompensasi.

Keterbatasan dan Saran untuk Penelitian Selanjutnya

Keterbatasan dalam penelitian ini adalah sebagai berikut :

1. Penelitian ini tidak mengukur level peningkatan pengetahuan mengenai BSC yang diberikan kepada partisipan sebelum dan sesudah pemberian pengetahuan BSC.
2. Penelitian ini belum menggunakan pendekatan-pendekatan lain, seperti pendekatan akuntabilitas (Libby, Salterio & Webb, 2004) dan pendekatan pelatihan (Dilla & Steinbart, 2005) yang dapat mengurangi bias pengukuran umum dalam penilaian kinerja menggunakan BSC. akuntabilitas.

Saran dalam penelitian ini adalah sebagai berikut:

1. Penelitian selanjutnya disarankan untuk mengukur level peningkatan pengetahuan mengenai BSC dengan memberikan pertanyaan-pertanyaan terkait teori dan penerapan BSC.
2. Penelitian selanjutnya disarankan dapat menggunakan pendekatan-pendekatan

lainnya, seperti pendekatan akuntabilitas (Libby, Salterio & Webb, 2004) dan pendekatan pelatihan (Dilla & Steinbart, 2005) untuk mengurangi bias pengukuran umum dalam penilaian kinerja menggunakan BSC.

DAFTAR PUSTAKA

- Aryani, Y. A. (2009). The Effect Of Fairness Perception Of Performance Measurement In The Balanced Scorecard Enviroment. *Master's Degree Thesis, Victoria University, Melbourne*.
- Banker, R. D., Chang, H., & Pizzini, M. J. (2004). *The Balanced Scorecard: Judgmental Effects of Perfomance Measures Linked to Strategy. The Accounting Review*, 79(1), 1-23.
- Bawono, I. R., Halim, A., & Lord, B. (2012, November). *Public Sector Performance Measurement and Budget Allocation: an Indonesian Experiment*.
- Bazerman, M. H. & Moore, D. A. (2009). *Judgment in Managerial Decision Making* (7th ed.). New Jersey: John Wiley & Sons.
- Bone, H. & Sholihin, M. (2012). Pengaruh Perspektif Dan Jenis Ukuran Dalam Balanced Scorecard Terhadap Evaluasi Kinerja. *Thesis, Universitas Gadjah Mada*.
- Bonner, S. E. (1990). Experience Effects in Auditing: The Role of Task-Spesific Knowledge. *The Accounting Review*, 65 (1), 72-92.
- Bonner, S. E., & Lewis, B. L. (1990). Determinants of Auditor Expertise. *Journal of Accounting and Auditing*, 28, 1-20.
- Dilla, W. N., & Steinbart, P. J. (2005). Relative Weighting of Common and Unique Measures by Knowledgeable Decision Makers. *Behavioral Research in Accounting*, 17 (1), 43-53.
- Frigo, M. L., & Krumwiede, K. R. (2000). *The Balanced Scorecard: A Winning Performance Measurement System. Strategic Finance*, 81(7), 50-54.
- Gagne, M. L., Hollister, J., Tully, G. J. (2006). Using the Balanced Scorecard: Both Common And Unique Measure Are Informative. *Journal of Applied Business Research*, 22, 147-160.
- Ghozali, Imam. (2001). *Aplikasi Analisis Multivariate dengan program SPSS*. Semarang: BP Universitas Diponegoro.
- Ghozali, Imam. (2007). *Aplikasi Analisis Multivariat dengan Program SPSS*. Semarang: BP Universitas Diponegoro.
- Ghozali, Imam. (2009). *Ekonometrika. Teori, Konsep dan Aplikasi dengan SPSS 17*. Semarang: BP Universitas Diponegoro.

- Gibbs, M. J, Merchant, K. A., Van der Stede, W. A., & Vargus, M. E. (2005). The Benefits of Evaluating Performance Subjectively. *Performance Improvement*, 44 (5), 26-32
- Grevinga, K. H. M. (2013). Common Measure Bias in the Balanced Scorecard: an Experiment with Undergraduate Students. *Master's Degree Thesis, University of Twente, Netherlands*.
- Gujarati, D. N. (1999). *Essential of Econometrics* (2th ed.). New York: McGraw-Hill.
- Gujarati, D. N. (2003). *Basic Econometrics* (4th ed.). New York: McGraw-Hill.
- Holmstrom, B., & Milgrom, P. (1991). *Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design*. *Journal of Law, Economics, & Organization*, 7, 24-52.
- Humphreys, K. A., & Trotman, K. T. (2011). *The Balanced Scorecard: The Effect of Strategy Information on Performance Evaluation Judgments*. *Journal of Management Accounting Research*, 23, 81-98.
- Indriantoro, N., & Supomo, iB. (1999). *Metode Penelitian Bisnis untuk Akuntansi dan Manajemen*. Yogyakarta: BPFE.
- Ittner, C. D., Larcker, D. F., & Meyer, M. W. (2003). Subjectivity and the Weighting of Performance Measures: Evidence from a Balanced Scorecard. *The Accounting Review*, 78 (3), 725-758.
- Kaplan, R. S., & Norton, D. P. (1992). The Balanced Scorecard - Measures that Drive Performance. *Harvard Business Review*, 70 (1), 71-79.
- Kaplan, R. S., & Norton, D. P. (1996). Linking the Balanced Scorecard to Strategy. *California Management Review*, 39 (1), 53-79.
- Kaplan, R. S., & Norton, D. P. (2001). The Strategy - Focused Organization: How Balanced Scorecard Companies Thrive in the New Business Environment. *Harvard Business School Press*.
- Kaplan, R. S., & Norton, D. P. (2001a). Transforming the Balanced Scorecard from Performance Measurement to Strategic Management (Part I). *Accounting Horizons*, 15 (1), 87-104.
- Kaplan, R. S., & Norton, D.P. (2001b). Transforming the Balanced Scorecard from Performance Measurement to Strategic Management (Part II). *Accounting Horizons*, 15 (2), 147-160.
- Kaskey, V. L. (2008). The Balanced Scorecard: A Comparative Study of Accounting Education and Experience on Common Measure Bias and Trust in a Balanced Scorecard. *PhD Dissertation. School of Business and Technology, Capella University, USA*.
- Krumwiede, K. R., Swain, M. R., Thornock, T. A., & Eggett, D. L. (2013). *The Effects of Task Outcome Feedback and Broad Domain Evaluation Experience on the Use of Unique Scorecard Measures*. *Advances in Accounting, incorporating Advances in International Accounting*, 29, 205-217.
- Libby, T., Salterio, S., & Webb, A. (2002). The Balanced Scorecard: The Effect of Assurance and Process Accountability on Managerial Judgment.
- Liedtka, S. L., Church, B. K., & Ray, M. R. (2008). Performance Variability, Ambiguity Intolerance, and Balanced Scorecard-Based Performance Assessments. *Behavioral Research in Accounting*, 20 (2), 73-88.
- Lindberg, E., & Schonfeldt, S. N. (2008). The Balanced Scorecard at Skelleftea Municipality. *Bachelor Thesis, UMEA School of Business, Sweden*.
- Lipe, M. G., & Salterio, S. E. (2000). The Balanced Scorecard: Judgmental Effects of Common and Unique Performance Measures. *The Accounting Review*, 75 (3), 283-298.
- Malina, M. A. & Selto, F. H. (2001). Communicating and Controlling Strategy An Empirical Study for The effectiveness of The Balanced Scorecard. *Journal Management Accounting Research*, 13, 47-90.
- Miqdad, M. (2012). *Praktik Tata Kelola Perusahaan (Corporate Governance) dan Usefulness Informasi Akuntansi (Telaah Teoritis dan Empiris)*. *Jurnal Manajemen dan Kewirausahaan*, 14 (2), 145-153.
- Niven, P. R. (2006). *Balanced Scorecard Step-by-Step Maximizing Performance and Maintaining Results* (2th ed.). New Jersey: John Wiley & Sons.
- Payne, J., J.Bettman & E. Johnson. (1993). *The Adaptive Decision Maker*. Cambridge, U.K.: Cambridge University Press.
- Rigby, D., Bilodeau, B. (2013). *Management Tools & Trends 2013*. *Bain & co*.
- Roberts, M. L., Albright, T. L., & Hibbets, A. R. (2004). Debiasing Balanced Scorecard Evaluations. *Behavioral Research in Accounting*, 16 (1), 75-88.
- Serfas, S. (2011). *Cognitive Biases in the Capital Investment Context: Theoretical Consideration and Empirical Experiments on Violations of Normative Rationality*. *Dissertation, Chemnitz University of Technology, Germany*. Retrieved from <http://books.google.co.id/books?id=i7OJWje1I>

[QC&printsec=frontcover#v=onepage&q&f=false](#)

- Tversky, A., & Kahneman, D. (1973). Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology*, 5, 207-232.
- Tversky, A., & Kahneman, D. (1974). Judgment Under Uncertainty: Heuristics and Biases. *Science*, 185, 1124-1131.